Staggered Cryst

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/88987 A2

(51) International Patent Classification[7]: H01L 27/112

(21) International Application Number: PCT/US01/14134

(22) International Filing Date: 1 May 2001 (01.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/204,467 16 May 2000 (16.05.2000) US
09/718,771 22 November 2000 (22.11.2000) US

(71) Applicant: ADVANCED MICRO DEVICES, INC. [US/US]; One AMD Place, Mail Stop 68, Sunnyvale, CA 94088-3453 (US).

(72) Inventors: RANDOLPH, Mark, W.; 3637 Kendra Way, San Jose, CA 95130 (US). HOLLMER, Shane, Charles; 1964 Conifer Lane, San Jose, CA 95132 (US). CHEN, Pau-Ling; 12947 Arroyo de Arguello, Saratoga, CA 95070 (US). FASTOW, Richard, M.; Apartment #2, 10203 Parkwood Drive, Cupertino, CA 95014 (US).

(74) Agent: RODDY, Richard, J.; Advanced Micro Devices, Inc., One AMD Place, Mail Stop 68, Sunnyvale, CA 94088-3453 (US).

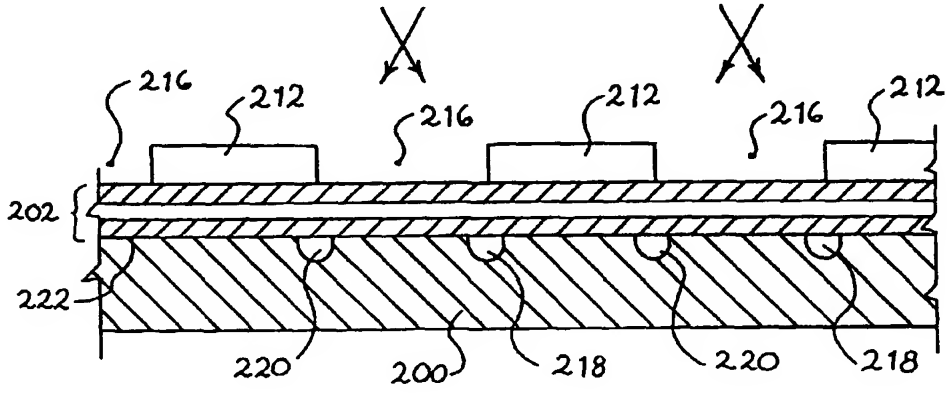
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: STAGGERED BITLINE STRAPPING OF A NON-VOLATILE MEMORY CELL



(57) Abstract: An array of memory cells includes a plurality of memory cells interconnected via a grid of M wordlines and M bitlines (224), wherein M = 2,3,4,5,..., wherein each of the M bitlines (224) is buried. The array further includes a plurality of contacts (228), wherein each of the plurality of contacts (228) is formed every N wordlines, N = 1,2,3,..., wherein each of the plurality of contacts (228) overlies a gate (229) of a different one of the plurality of memory cells. A strap (230) connects one of the buried bitlines (241) to a gate (229) that underlies one of the plurality of contacts (228), and wherein contacts (228) overlying a first bit line (224) are staggered with respect to contacts (228) overlying a second bit line (224) that is adjacent to the first bit line (224).

WO 01/88987 A2

STAGGERED BITLINE STRAPPING OF A NON-VOLATILE MEMORY CELL

TECHNICAL FIELD

The present invention relates to the field of non-volatile memory devices. More particularly, the invention relates to a multi-bit flash electrically erasable programmable read only memory (EEPROM) cell with a bitline.

BACKGROUND ART

Memory devices for non-volatile storage of information are currently in widespread use today, being used in a myriad of applications. A few examples of non-volatile semiconductor memory include read only memory (ROM), programmable read only memory (PROM), erasable programmable read only memory (EPROM), electrically erasable programmable read only memory (EEPROM) and flash EEPROM.

Semiconductor EEPROM devices involve more complex processing and testing procedures than ROM, but have the advantage of electrical programming and erasing. Using EEPROM devices in circuitry permits in-circuit erasing and reprogramming of the device, a feat not possible with conventional EPROM memory. Flash EEPROMs are similar to EEPROMs in that memory cells can be programmed (i.e., written) and erased electrically but with the additional ability of erasing all memory cells at once, hence the term flash EEPROM.

An example of a single transistor Oxide-Nitrogen-Oxide (ONO) EEPROM device is disclosed in the technical article entitled "A True Single-Transistor Oxide-Nitride-Oxide EEPROM Device," T.Y. Chan, K.K. Young and Chenming Hu, IEEE Electron Device Letters, March 1987. The memory cell is programmed by hot electron injection and the injected charges are stored in the oxide-nitride-oxide (ONO) layer of the device. Other examples of ONO EEPROM devices are disclosed in U.S. Patents Nos. 5,635,415; 5,768,192 and PCT patent application publication WO 99/07000, the contents of each reference are hereby incorporated herein by reference.

In the case of known NROM devices, such as schematically shown in FIG. 1, an NROM cell 100 included a grid of polygates or word lines 102 and buried bitlines 104. The bitlines 104 were formed in the N+ region of the substrate so that a higher density of bitlines can be formed that region versus when the bitlines were formed in a metal layer. Select transistors 106 were required to be placed every N or N/2 polygates 102, where N is the number of polygates between contacts 108. This in the past has required a select transistor 106 being required every 16 or 32 cells in order to reduce the bitline to cell resistance. The bitline resistance in the N+ region limits the number of cells between select transistors.

In the case of flash memory cells with a stacked gate, contacts associated with the cell must be spaced from the polysilicon of the gate. As feature sizes are reduced according to integrated circuit processes, smaller dimensions are required to achieve higher packing densities. Generally, contacts must be spaced apart from the stacked gate so alignment errors do not result in a shorting of the stacked gate with the source contact or the drain contact. The spacing between the c ntact and gate contributes to the overall size of the flash memory cell.

DISCLOSURE Of THE INVENTION

One aspect f the invention regards an array of memory cells includes a plurality of memory cells interconnected via a grid of M wordlines and M bitlines, wherein M = 2, 3, 4, 5, . . ., wherein each of the M bitlines is buried. The array further includes a plurality of contacts, wherein each of the plurality of contacts is formed every N wordlines, N = 1, 2, 3, . . ., wherein each of the plurality of contacts overlies a gate of a different one of the plurality of memory cells. A strap connects one of the buried bitlines to a gate that underlies one of the plurality of contacts, and wherein contacts overlying a first bit line are staggered with respect to contacts overlying a second bit line that is adjacent to the first bit line.

The above aspect of the present invention provides the advantage of reducing the source plus drain resistance per cell.

The above aspect of the present invention provides the advantage of eliminating the need for select transistors and reducing the total size of an array.

The present invention, together with attendant objects and advantages, will be best understood with reference to the detailed description below in connection with the attached drawings.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a top cross-sectional view of a prior art NROM cell with a buried bit line;

FIGS. 2-6 illustrate side cross-sectional views of processing steps to form an embodiment of the present invention;

FIG. 7 illustrates a side cross-sectional view of a two bit flash EEPROM cell constructed in accordance with an embodiment of the present invention utilizing the process of FIGS. 2-6; and

FIG. 8 illustrates a top cross-sectional view of the two bit flash EEPROM cell of FIG. 7.

MODE(S) FOR CARRYING OUT THE INVENTION AND INDUSTRIAL APPLICABILITY

Non-volatile memory designers have taken advantage of the ability of silicon nitride to store charge in localized regions and have designed memory circuits that utilize two regions of stored charge within the ONO layer. This type of non-volatile memory device is known as a two-bit EEPROM. The two-bit EEPROM is capable of storing twice as much information as a conventional EEPROM in a memory array of equal size. A left and right bit is stored in physically different areas of the silicon nitride layer, near left and right regions of each memory cell. Programming methods are then used that enable two-bits to be programmed and read simultaneously. The two-bits of the memory cell can be individually erased by applying suitable erase voltages to the gate and to either the source or drain regions.

Two bit memory cells are typically accessed by buried bit-lines formed in a semiconductor substrate. A bit-line oxide layer is formed over the buried bit-line prior to forming a central gate electrode.

Shown in FIG. 2, in cross-section is a portion of a semiconductor substrate 200 having already undergone several processing steps. An ONO layer 202 overlies the semiconductor substrate 200 and includes a first oxide layer 206, a second oxide layer 208 and a silicon nitride layer 210 sandwiched between the first oxide layer 206 and the second oxide layer 208.

As shown in FIG. 3, a resist layer 212 is formed to overly the ONO layer 202. Resist layer 212 can be one of a number of different types of resist, including optical photoresist responsive to visible and near UV

light, deep UV resist and the like. Alternatively, resist layer 212 can be an inorganic resist layer, an X-ray resist layer and the like. In a preferred embodiment, resist layer 212 is a N volak resin photoresist material.

Resist layer 212 is exposed to radiation of the appropriate wavelength and developed to form a resist pattern overlying ONO layer 202, as illustrated in FIG. 3. Resist pattern 212 is formed to have a predetermined geometric configuration for the fabrication of buried bit-line regions in semiconductor substrate 200. Resist pattern 212 allows for exposing selected regions 216 of semiconductor substrate 200. Once resist pattern 212 is formed, an implantation process is carried out to form pocket regions 218, 220 in semiconductor substrate 200. Pocket regions 218, 220 are preferably formed by an angled ion implant process in which semiconductor substrate 200 is held at an angle of about 7° to about 60°, typically 30° to 45°, with respect to normal during the ion implantation process. The angled ion implant process forms pocket regions 218, 220 in semiconductor substrate 200 in locations that partially underlie a portion of resist pattern 212. In a preferred embodiment, a p-type dopant, such as boron, is ion implanted into semiconductor substrate 200 to form pocket regions 218, 220. During the ion implantation process, the boron ions penetrate ONO layer 202 and enter semiconductor substrate 200 at an angle sufficient to create a boron pocket region that extends partially beneath resist pattern 212.

Referring to FIG. 4, after forming the pocket regions 218, 220, portions of ONO layer 202 exposed by resist pattern 212 are etched to expose principal surface 222 of semiconductor substrate 200. Preferably, resist pattern 212 is used as an etching mask, such that the etching process exposes principal surface 212 in selected regions 216 defined by resist mask 212. In a preferred embodiment, ONO layer 202 is anisotropically etched, such that ONO layer 202 and resist pattern 212 have continuous, substantially vertical sidewalls.

Once the etching process is complete, preferably an ion implantation process is carried out to form a buried bit-line region 224 in selected region 216 of semiconductor substrate 200. Preferably, an n-type dopant, such as arsenic, is ion implanted at an angle of incidence substantially normal to principal surface 222 of semiconductor substrate 200. Preferably, buried bit-line region 224 is formed by the ion implantation of arsenic using a dose of about 3 x 1015 to about 5 x 1015 ions per square centimeter. The ion implantation energy is selected so as to form buried bit-line region 224 to a selected junction depth in semiconductor substrate 200. Preferably, the ion implantation energy is of sufficient magnitude, such that the junction depth of buried bit-line region 224 is greater than the junction depth of pocket regions 218, 220. As used herein, the term "junction depth" refers to the distance from the surface of the substrate to the deepest point of formation of a p/n junction associated with the implanted region within the substrate.

Those skilled in the art will recognize that other methods for forming the memory cell arrays are possible. For example, the order of formation of the pocket regions 218, 220 and the buried bit-line region 224 can be reversed from that described above. In an alternative embodiment, before etching ONO layer 202, an implant process can be carried out to form bit-line region 224, followed by an angled implant process to form pocket regions 218, 220. In yet another alternative, ONO layer 202 can be etched before either implant process is carried out.

As illustrated in FIG. 5, the resist pattern 212 is removed and bit-line oxide regions 226 are formed. In a preferred embodiment, bit-line oxide layer 226 is formed by thermal oxidation of semiconductor substrate

200 using ONO layer 202 as an oxidation mask. ONO layer 202, having been previ usly patterned by the etching process described above, exposes selected regions 216 of semiconductor substrate 200. During the oxidation process, the patterned portions of ONO layer 202 prevent the xidation of semiconductor substrate 200 in region underlying ONO layer 202. Accordingly, bit-line oxide layers 226 are confined to selected regions 216 of semiconductor substrate 200. Upon completion of the oxidation process, bit-line layers 226 overly buried bit-line regions 224 in semiconductor substrate 200.

In addition to the layers 226, control gate electrode contacts/electrodes 228 are formed over the floating gate electrodes 229 by depositing a layer of polycrystalline silicon by a CVD process, followed by patterning and etching to form thin control-gate lines overlying the substrate 200. As shown in FIG. 6, the electrode 228 overlies the layers 226 and bit line oxide regions 224.

Once the above-described process is complete, a two bit flash EEPROM cell is formed as shown in FIG. 6. The flash EEPROM memory cell includes an N+ type substrate 200 having two buried PN junctions, one being between the source pocket 218 and substrate 200, termed the left junction and the other being between the drain pocket 220 and the substrate 200, termed the right junction. Above the channel 230 is an oxide layer 206 made of silicon dioxide. The oxide layer 206 has a thickness that is less than or equal to 60 Angstroms, and which forms an electrical isolation layer over the channel.

On top of the oxide layer 206 is a charge trapping layer 210 that has a thickness ranging from approximately 20 to 100 Angstroms and preferably is comprised of silicon nitride, Si3N4. The hot electrons are trapped as they are injected into the charge trapping layer so that the charge trapping layer serves as the memory retention layer.

The thickness of layer 210 is chosen to be in excess of approximately 50 Angstroms to prevent electrons from tunneling through the layer 206 and leaving charge trapping layer 210 during the operation of the cell. Thus, the lifetime of the cell of this invention is greatly extended relative to prior art NMOS devices. The memory cell is capable of storing two bits of data, a right bit and a left bit.

It is important to note that the two-bit memory cell is a symmetrical device. For example, the left junction serves as the source terminal and the right junction serves as the drain terminal for the right bit. Similarly, for the left bit, the right junction serves as the source terminal and the left junction serves as the drain terminal. Thus, the terms left, or first junction and right or second junction are used herein rather than source and drain. When the distinction between left and right bits is not crucial to the particular discussion, the terms source and drain are utilized. However, it should be understood that the source and drain terminals for the second bit are reversed compared to the source and drain terminals for the first bit.

A layer of silicon dioxide 208 is formed over the charge trapping layer, (i.e., silicon nitride layer), and has a thickness that ranges between approximately 60 to 100 Angstroms. The silicon dioxide layer 208 functions to electrically isolate a conductive gate 228 formed over the silicon dioxide layer 208 from charge trapping layer 210. The thickness of gate 228 is approximately 4,000 Angstroms. Gate 228 is constructed from an N-type material, such as polycrystalline silicon that is typically heavily doped with an N-type impurity such as phosphorous in the 1019 to 1020 atom/cc range.

As shown in the enlarged cross-sectional schematic view of FIG. 7, polysilicon straps 231 can be made concurrently with or with ut bitlines 226 and are used to connect each buried bitline 224 to the overlying gate electrode 228. As sh wn in FIG. 8, the bitlines 224 are continuous, uniform and unbroken. For the contacts 228 associated with a bit line 224, the contacts 228 are separated by a distance D from one another by a constant number of cells. The separation between contacts 228 is the same for each bitline. The contacts 228 are staggered or offset so that contacts 228 on adjacent bitlines are not adjacent to each other and where each contact 228 overlies a buried bitline 224. Preferably, the offset distance OD is equal to one half of the separation distance D. The contacts 228 are positioned along wordlines associated with polygates 229. By having a continuous bitline with staggered contacts the source plus drain resistance per cell can be reduced while eliminating the need for select transistors. Note that the resistance per cell may not be uniform. The elimination of select transistors reduces the total size of the memory array when compared with the array of FIG. 1. Please note that while FIG. 8 shows a portion of an M X M memory array where M= 10. The above principles can also be applied for when M= 2, 3, . . ., etc.

It is important to note that when a semiconductor device is scaled, the channel lengths become shorter and short channel effects take hold. Thus, in the two bit memory cell, because each bit is stored in different areas of the transistor, short channel effects may become prevalent sooner than in the case of the single bit transistor. In order to retain the usable range of drain voltage, the two-bit transistor may need to be scaled by a smaller factor.

The foregoing description is provided to illustrate the invention, and is not to be construed as a limitation. Numerous additions, substitutions and other changes can be made to the invention without departing from its scope as set forth in the appended claims.

CLAIM

WE CLAIM:

1. An array of memory cells comprising:

a plurality of memory cells interconnected via a grid of M wordlines and M bitlines (224), wherein M = 2, 3, 4, 5, . . ., wherein each of said M bitlines (224) is buried;

a plurality of contacts (228), wherein each of said plurality of contacts (228) is formed every N wordlines, N = 1, 2, 3, . . ., wherein each of said plurality of contacts (228) overlies a gate (229) of a different one of said plurality of memory cells;

a strap (231) connecting one of said buried bitlines (224) to a gate (229) that underlies one of said plurality of contacts (228); and

wherein contacts (228) overlying a first bit line (224) are staggered with respect to contacts (228) overlying a second bit line (224) that is adjacent to said first bit line (224).

2. The array of memory cells of claim 1, wherein said contacts (228) overlying said first bit line (224) are equally spaced from one another by a distance D.

3. The array of memory cells of claim 2, wherein said contacts (228) overlying said second bit line (224) are equally spaced from one another by a distance D.

4. The array of memory cells of claim 2 or claim 3, wherein said contacts (228) overlying said first bit line (224) are staggered by a distance D/2 with respect to said contacts (228) overlying said second bit line (224).

5. The array of memory cells of claim 1, wherein the bitlines (224) are continuous, uniform and unbroken.

6. The array of memory cells of claim 1, wherein each of said plurality of contacts (228) overlies a buried bitline (224).

7. The array of memory cells of claim 5, wherein each of said plurality of contacts (228) overlies a buried bitline (224).

8. The array of memory cells of claim 1, wherein each of said plurality of memory cells comprises:

a substrate (200) that comprises a first region (218) and a second region (220) with a channel (230) therebetween and a gate (229) above said channel (230),

a charge trapping region (210) that contains a first amount of charge, and

a layer (206) positioned between said channel (230) and said charge trapping region (210), wherein said layer (206) has a thickness such that said first amount of charge is prevented from directly tunneling into said layer (206).

9. The array of memory cells of claim 8, wherein each of said plurality of memory cells comprises an EEPROM memory cell.

10. The array of memory cells of claim 8, wherein each of said plurality of memory cells comprises a two bit memory cell.

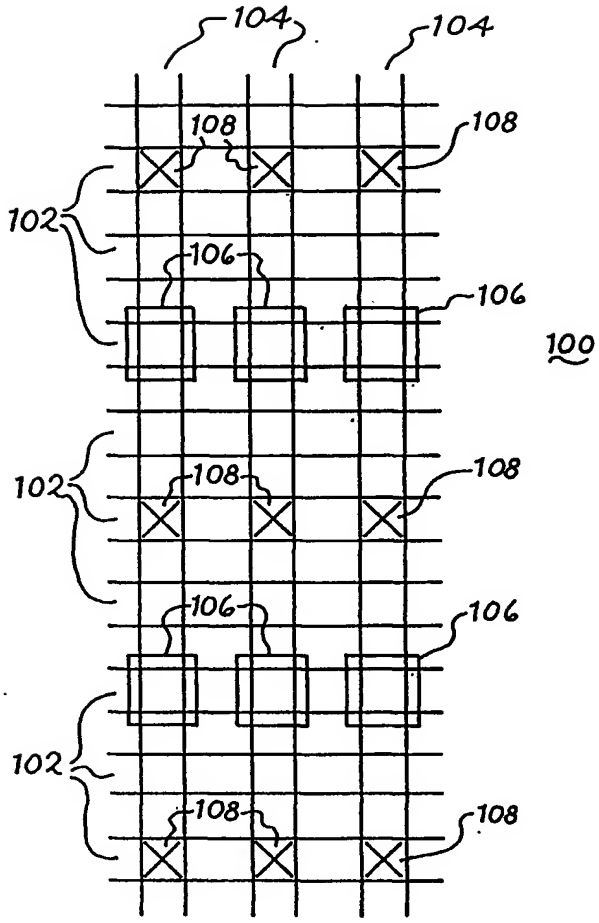


Fig. 1

Fig. 2

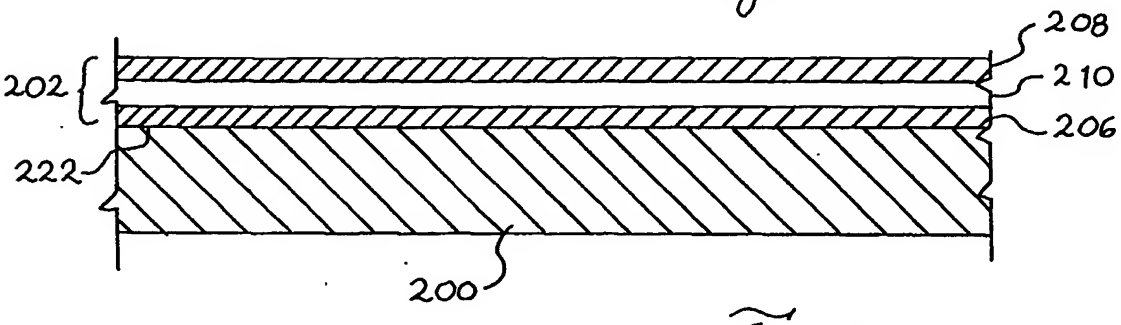


Fig. 3

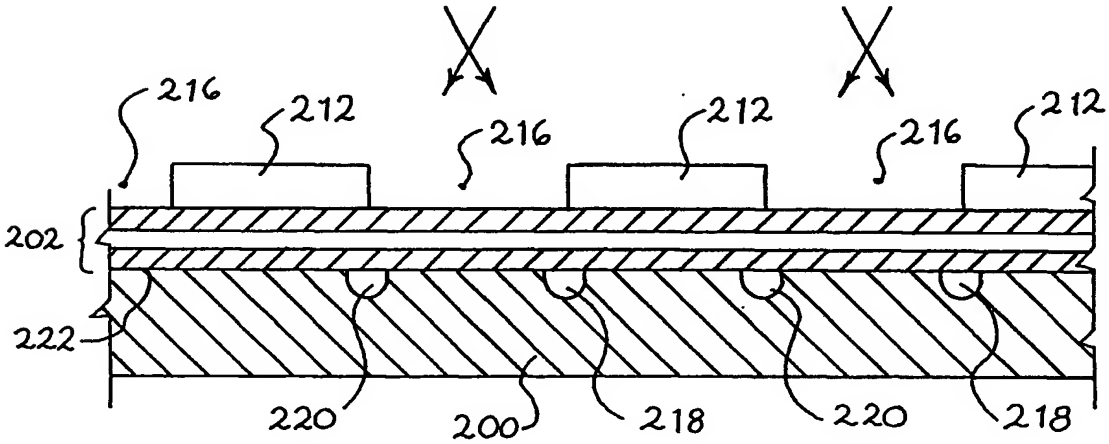


Fig. 4

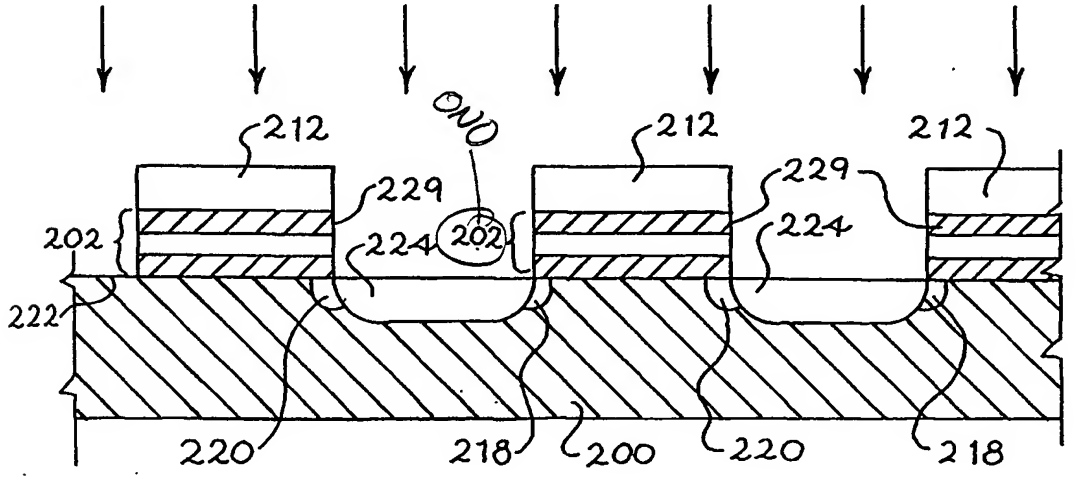


Fig. 5

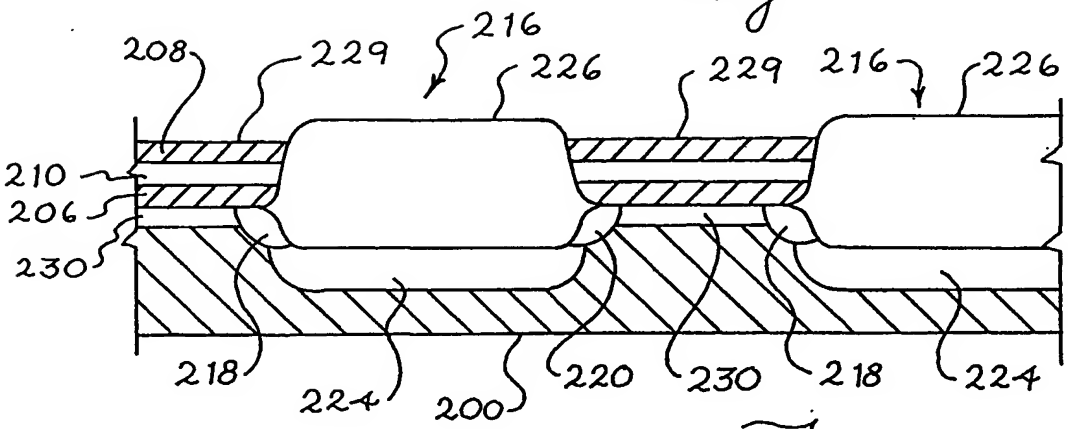


Fig. 6

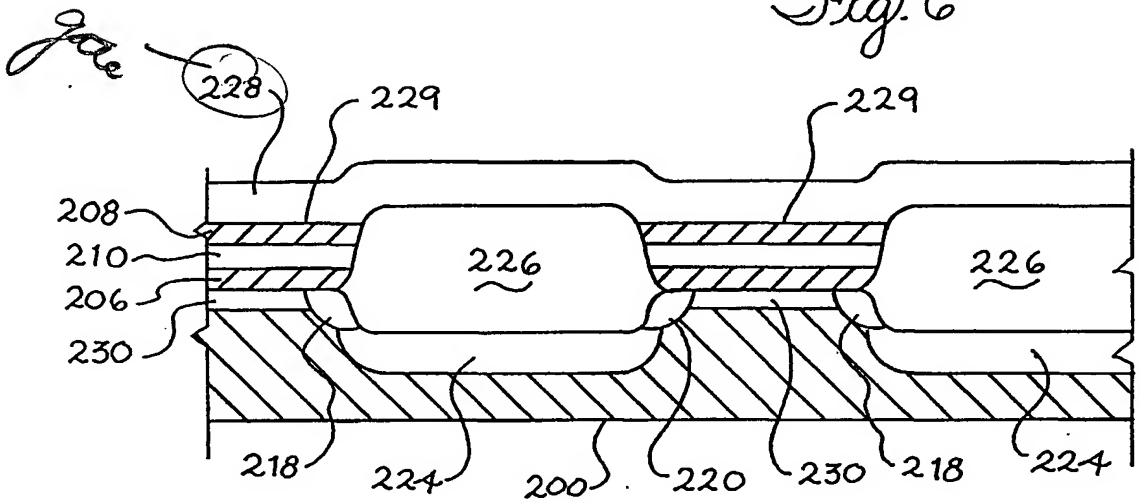
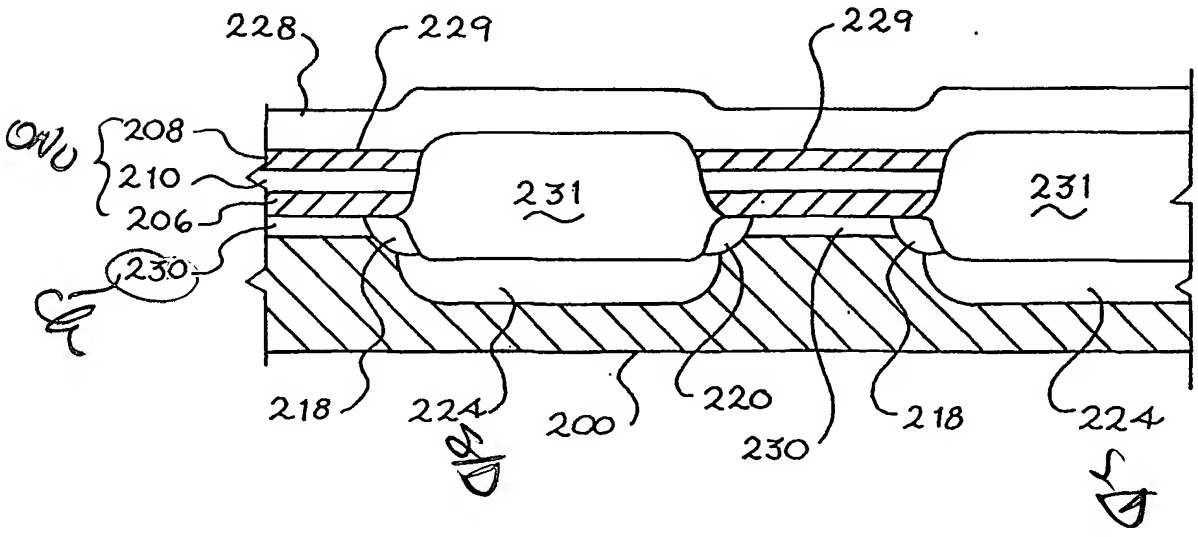


Fig. 7



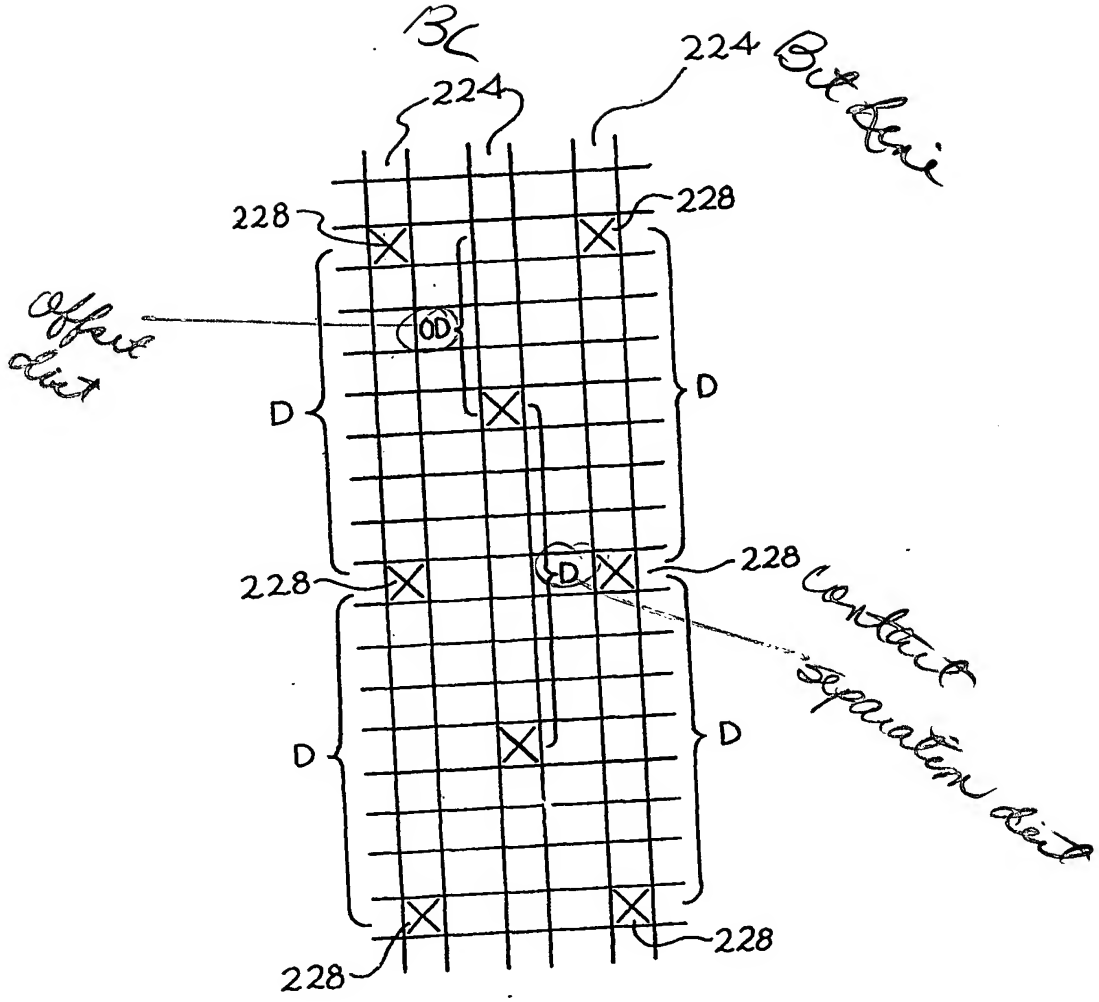

BC
224
224
228
228
228
228
228
228
OD
D
D
D
D
Offset dist
Contact Separation dist

Fig. 8